iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.51624

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

A Research on YouTube Spam Comments Detection and Deletion

Darshan Bhavsar[1], Stephen Dcruz[2], Abhishek Chandekar[3], Mihir Chaudhari[4], Prof .Chetana Patil[5]
[1, 2, 3, 4, 5]*Dhole Patil College of Engineering Pune, India*

Abstract: ***YouTube is a popular social networking platform that allows users to upload, watch and comment on videos. However, YouTube also attracts spammers who post unwanted or abusive comments on videos, which can affect the quality and credibility of the content and platform. In this paper, we propose a novel approach to detect and delete YouTube spam comments using natural language processing (NLP) algorithms such as Rapid Fuzz and Levenshtein distance. Rapid Fuzz is a Python library that provides fast and accurate fuzzy string matching, while Levenshtein distance is a metric that measures the minimum number of edits required to transform one string into another. We use these algorithms to compare the similarity of comments and identify duplicates, which are a common form of spam. We also use the YouTube API to access and delete a channel's spam comments. We evaluate our approach on comments of YouTube Video and show that it can effectively detect and delete spam comments with high accuracy and efficiency.***
Keywords: YouTube, Spam Comments Detection, Levenshtein Distance, Rapid Fuzz, Natural Language Processing.

I. INTRODUCTION

YouTube is a popular social networking platform that allows users to upload, watch and comment on videos. However, YouTube also attracts spammers who post unwanted or abusive comments on videos, which can affect the quality and credibility of the content. Spam comments can be defined as comments that are irrelevant, repetitive, misleading, or malicious. They can also contain links to harmful or inappropriate websites, or promote products or services without the consent of the video owner. Spam comments can harm the reputation of the video owner, discourage genuine viewers from engaging with the content, and violate the YouTube community guidelines. Therefore, it is important to detect and delete spam comments on YouTube videos. However, this task is not easy due to the large volume and variety of comments posted on YouTube every day. According to YouTube statistics, more than 500 hours of video are uploaded every minute, and more than 2 billion logged-in users visit YouTube each month. Moreover, spam comments can be written in different languages, use different formats and styles, and employ various techniques to evade detection made by YouTube itself. YouTube spam comments can be classified into different types based on their content and purpose. Spam comments include self-promotion, phishing, malware distribution, and hate speech.

In this paper, we propose an approach to detect and delete YouTube spam comments using natural language processing (NLP) algorithms such as Rapid Fuzz and Levenshtein distance. NLP is a subfield of artificial intelligence that examines how computers and human languages interact. Rapid Fuzz is a Python library that provides fast and accurate fuzzy string matching, while Levenshtein distance is a metric that measures the minimum number of edits required to transform one string into another. We use these algorithms to compare the similarity of comments and identify duplicates, which are a common form of spam. We use the YouTube Data API V3 to access and delete the spam comments on a given channel by using a token file for User Authentication which can be generated by creating a project on the google cloud console. User Must allow necessary permissions to our application so that, it can modify the comments of user's video by accessing YouTube data using the SSL security method from the browser. Without a token file, the application will not have access to your YouTube Channel due to security reasons applied by YouTube.

II. LITERATURE SURVEY

Identifying spam users and content focuses on a variety of areas. We reviewed the related work on YouTube spam comment detection and deletion by focusing on the definition and characteristics of YouTube spam comments, the existing methods and techniques for detecting and deleting them, and the challenges and limitations of the current research. YouTube spam comments are comments that are irrelevant, repetitive, misleading, or malicious. They can also contain links to harmful or inappropriate websites, or promote products or services without the consent of the video owner. Spam comments can harm the reputation of the video owner, discourage genuine viewers from engaging with the content, and violate the YouTube community guidelines.

©IJRASET: All Rights are Reserved

Spammers target YouTube with low-quality content or advertisements as it becomes more and more popular as a platform for sharing videos. The number of spammers harming the YouTube community is growing, making it interesting to conduct a study on how to identify them. Therefore, we carry out the research on identifying spam Comments on YouTube Platform.

YouTube spam comments can be classified into different types based on their content and purpose. YouTube spam comments consist of four categories: self-promotion, phishing, malware distribution, and hate speech. Self-promotion comments are those that advertise or endorse a product, service, website, or channel without permission. Phishing comments are those that attempt to trick users into revealing their personal or financial information. Malware distribution comments are those that contain links to malicious websites or files that can infect a user's device. Hate speech comments are those that express hatred or discrimination against a person or group based on their identity or characteristics. YouTube spam comments can also vary in their format and style, where Spam bots are made having emojis and phone numbers written in a less readable format which can be detected by our proposed system. Spam comment bots are typically created using automated software or scripts that can be programmed to perform a specific set of actions, such as posting comments on YouTube videos.

The process of creating a spam comment bot usually involves the following steps:

1) Creating a fake YouTube account : The bot creator will create a fake YouTube account, which will be used to post spam comments.
2) Extracting video URLs : The bot will scan the YouTube website for videos related to a particular topic, and extract the URLs of these videos.
3) Generating comments : The bot will generate a set of comments based on a pre-defined set of keywords and phrases. These comments are typically designed to be generic and non-specific, so that they can be used on a wide range of videos.
4) Posting comments : The bot will use the fake YouTube account to post the generated comments on the videos that were previously identified. The comments may be posted in bulk, or spread out over time to avoid detection.

Overall, the process of creating a spam comment bot involves a combination of technical expertise and knowledge of YouTube's platform and algorithms. It is important to note that spamming violates YouTube's terms of service and can result in penalties, including account suspension or termination .

A. Spam Tools

Scrape Box, Xrumer, Comment Blaster , Tube Assist, and Auto Hot Key are some of the tools used by spammers to make spam comments. Auto Hot Key is a tool that uses '.ahk' scripts that map keyboard keys with software for automatic operations. We have used the Auto Hot key tool for testing our System to mimic the spammer techniques for Effective and Efficient spam detection.

We do not Promote any Spam tools and techniques and used them only for testing Purposes of our Application.

III. METHODOLOGY

Spam comments are a common problem on YouTube, as they can affect the quality and engagement of the comment section. Spam comments can be identified by various features, such as the content, the username, the number of likes, and the presence of emojis. In this paper, we propose a method for detecting and deleting spam comments using natural language processing techniques such as Levenshtein distance and the Rapid Fuzz algorithm.

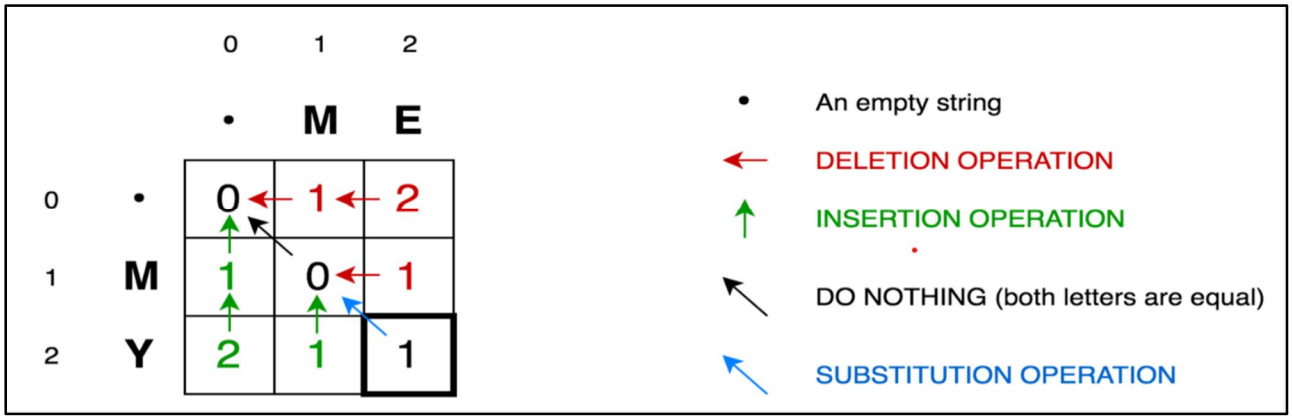


Fig 1. Levenshtein Matrix

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Fig 1. Shows the Levenshtein matrix used to calculate the distance between the strings. Levenshtein distance is a string metric that measures the minimum number of insertions, deletions, or substitutions required to transform one string into another. For example, the Levenshtein distance between “cat” and “bat” is 1, because one substitution is needed. Levenshtein distance can be used to compare the similarity between two comments or between a comment and a predefined spam template. A low Levenshtein distance indicates a high similarity and a possible spam comment.

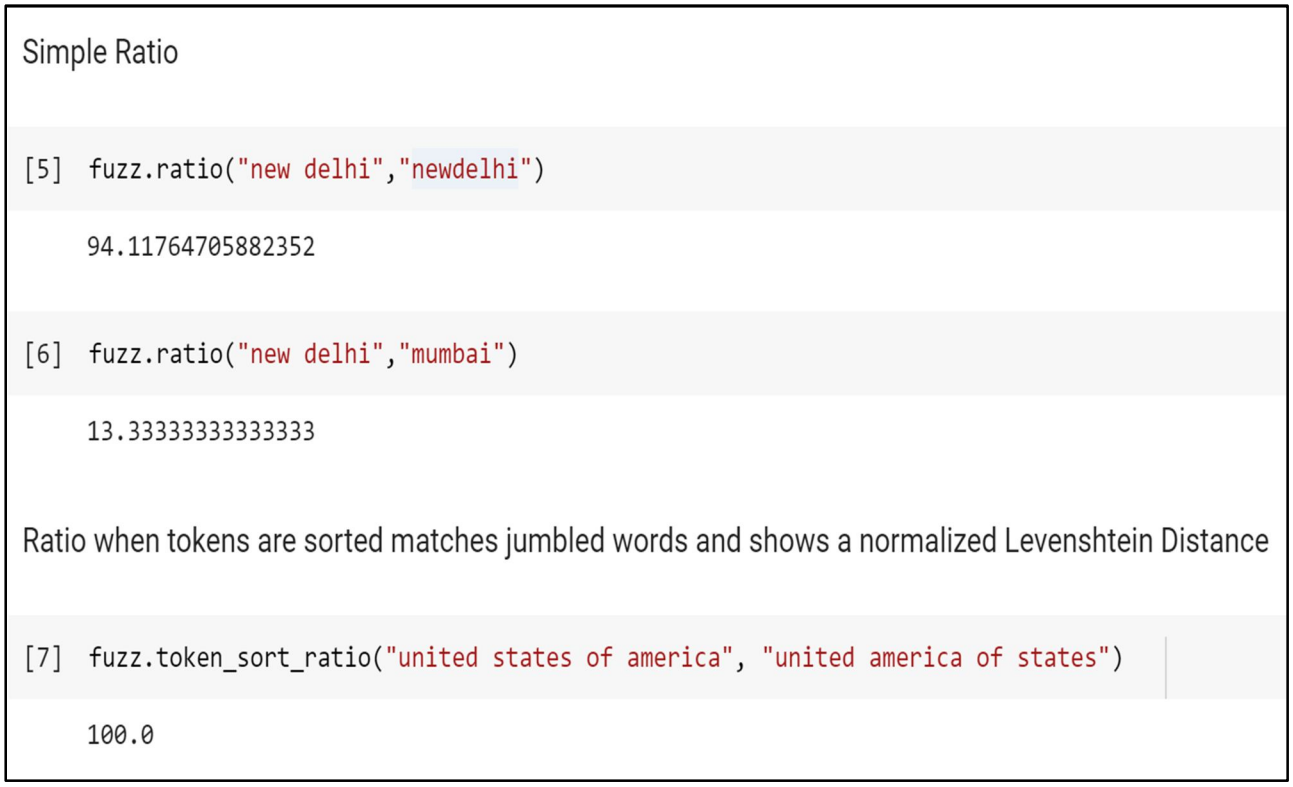


Fig 2. Rapid Fuzz

Fig 2 Shows Rapid Fuzz Algorithm which is a fast-string matching library for Python and C++ that implements various string metrics, including Levenshtein distance. Rapid Fuzz can also perform partial matching, token sorting, and token set matching, which can handle different word orders and repetitions. Rapid Fuzz can be used to efficiently compute the Levenshtein distance and other metrics between comments and spam templates.

We use the Levenshtein distance and Rapid Fuzz algorithm to detect spam comments that include emojis and numbers in usernames. Emojis and numbers are often used by spammers to attract attention and bypass filters. We pre-process the comments by removing punctuation marks and converting them to lowercase.

We then compare the usernames of the comments with a list of spam usernames that contain emojis and numbers using the Levenshtein distance and Rapid Fuzz algorithm. If the similarity score is above a certain threshold, we mark the comment as spam and delete it from the comment section. We evaluate our method on a dataset of YouTube comments from various popular videos. We compare our method with existing methods such as Support Vector Machine, K-Nearest Neighbour, and Ensemble Machine Learning Model. Our results show that our method achieves high accuracy, precision, recall, and F1-score in detecting and deleting spam comments that include emojis and numbers in usernames.

IV. SYSTEM DESIGN

Our Proposed System is a Windows Executable Program Made using Python Language, which can be used by A YouTube channel owner, or by a normal user allowing the user to detect and delete Spam Comments on their videos. Following Fig 3. Below Is the system Architecture Diagram for a clear understanding of the system, which shows the system flow from start to end.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

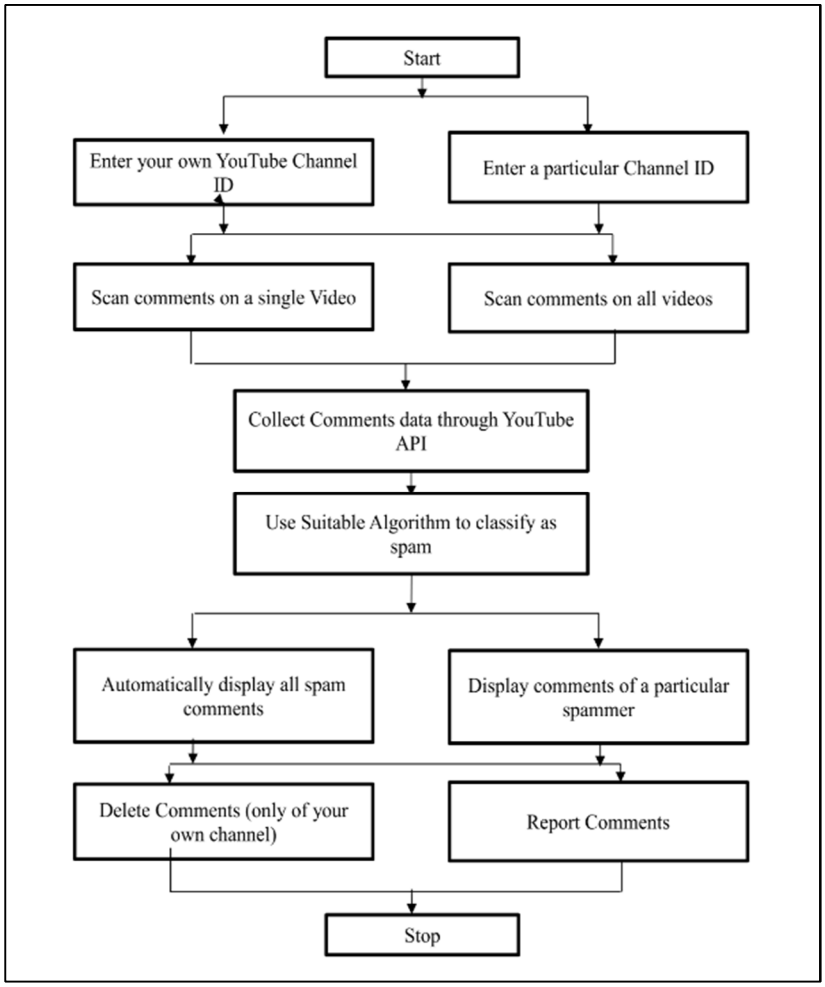


Fig 3. System Architecture

A. Initial Setup

First You need to create an Api Key from the Google cloud platform which will allow our program access to your YouTube channel account. Later you need to complete authentication with the Application to allow the Program to change your YouTube Video comment Section. You need to log in with your Google Account linked with your YouTube channel for detecting spam comments.

B. Our System Consists of 3 Scanning Modes

1) Scan specific video URL (Enter a video URL of User's channel or some other channel to scan its comments.
2) Scan a specific community post by its URL present in Address Bar.
3) Recover deleted comments using a log file (Only works on User's own channel).

The Specific URL option allows the User to put His / Her video URL link in the system by simply copying it from the address bar and pasting it in our application. Community post of a user can also be scanned for any spam comments using its URL Id. If the User has detected and deleted the spam comments and wishes to get them back, then it can be done using a log file saved during detection in the installation folder of the application, where comments can be recovered using YouTube's API method.

C. Spam Identification can be Done in 4 Ways

1) Auto Scan Mode: Automatically detects multiple spammer techniques.
2) Enter Spammer's channel ID(s) or link(s).
3) Scan usernames for the criteria you choose.
4) Scan comments for criteria you choose.

Automatic Scan mode scans the video comments and matches them with a filter Variables file present in our system which detects if the comment contains any spam emojis or words which are misleading people by redirecting them to telegram or Whatsapp bots by malicious links or phone numbers. The system also allows you to manually put a spammer's channel id if you wish to detect a specific Spam username. You can also enter a comment text eg. 'Click on this link or 'Message me on WhatsApp' or 'Dm me on Telegram' in our system, if the Auto Scan mode fails to detect these comments.

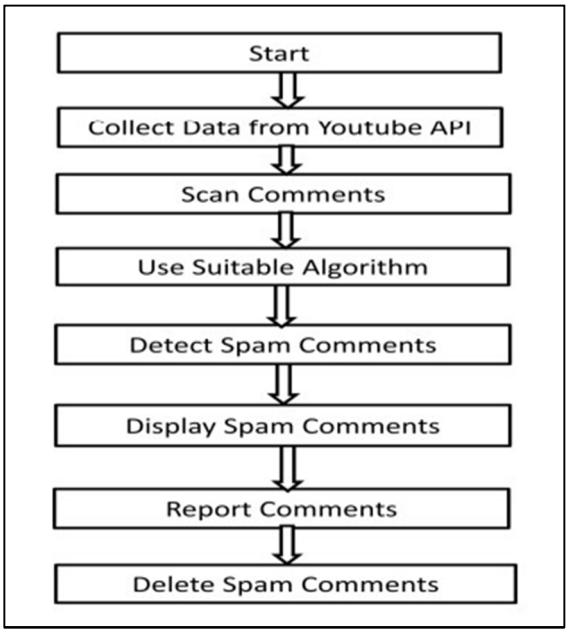


Fig 4. System Flow Diagram

Above Fig 4. Shows the flow of our system, where User Authentication Data is collected from YouTube API, then User enters a video or community post URL link in our System for detection where he/she can select various Identification methods for easy detection of spam comments on their channel. A user can delete the detected comments from the video or report these comments to YouTube.

The System can work on User's own Channel video where He/she can detect and delete spam comments, while it also has a 'Not Your Channel Mode' where it can detect spam comments on another channel which can be reported to YouTube all at once, thereby improving YouTube Algorithm to detect spam comments. YouTube data API fair Usage has a daily limit of 10,000 units per day, which allows comment modification. No of units used daily depends on the type of operations performed. The System Scans a Huge number of comments by using threads, where comments are divided into batches and executed sequentially for faster execution.

V. RESULTS

MODES	DETECTION RATE (%)
1) Auto Smart Mode	90
2) Enter Spammer's Channel ID	80
3) Scan Usernames for the criteria you choose	85
4) Scan Comment text for the criteria you choose	90

Fig 5. Scanning Modes

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

1)	Raj1@32Bot:	Hello dm me on WhatsApp
2)	Raj1@32Bot:	Hello dm me on WhatsApp
3)	Raj1@32Bot:	Hello dm me on WhatsApp
4)	Raj1@32Bot:	Hello dm me on WhatsApp
	Spammer Detected	
	REASON:	Multiple Duplicate Comments
	Direct Link:	//.wwwyoutube.com/watch?v=D7LPCMOXxSA&lc=Ugxyz3WZQ1V_lpp211
	Author Channel ID:	UC3QutT_Kkb3p3YkfgaOXaQ

Fig 6. Spam Comment Example

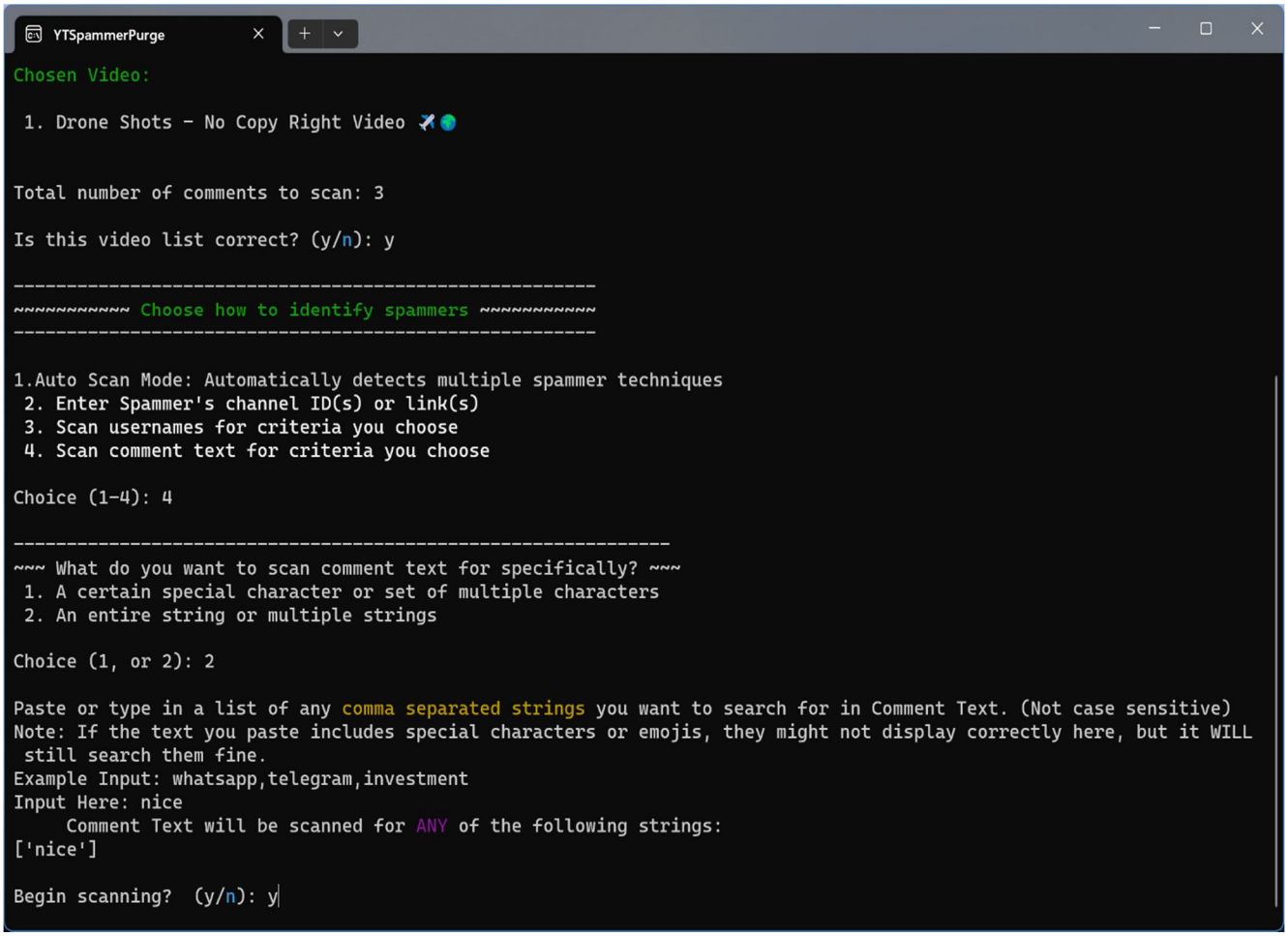


Fig 7. Screen Shot of Our Application displaying Scanning Modes

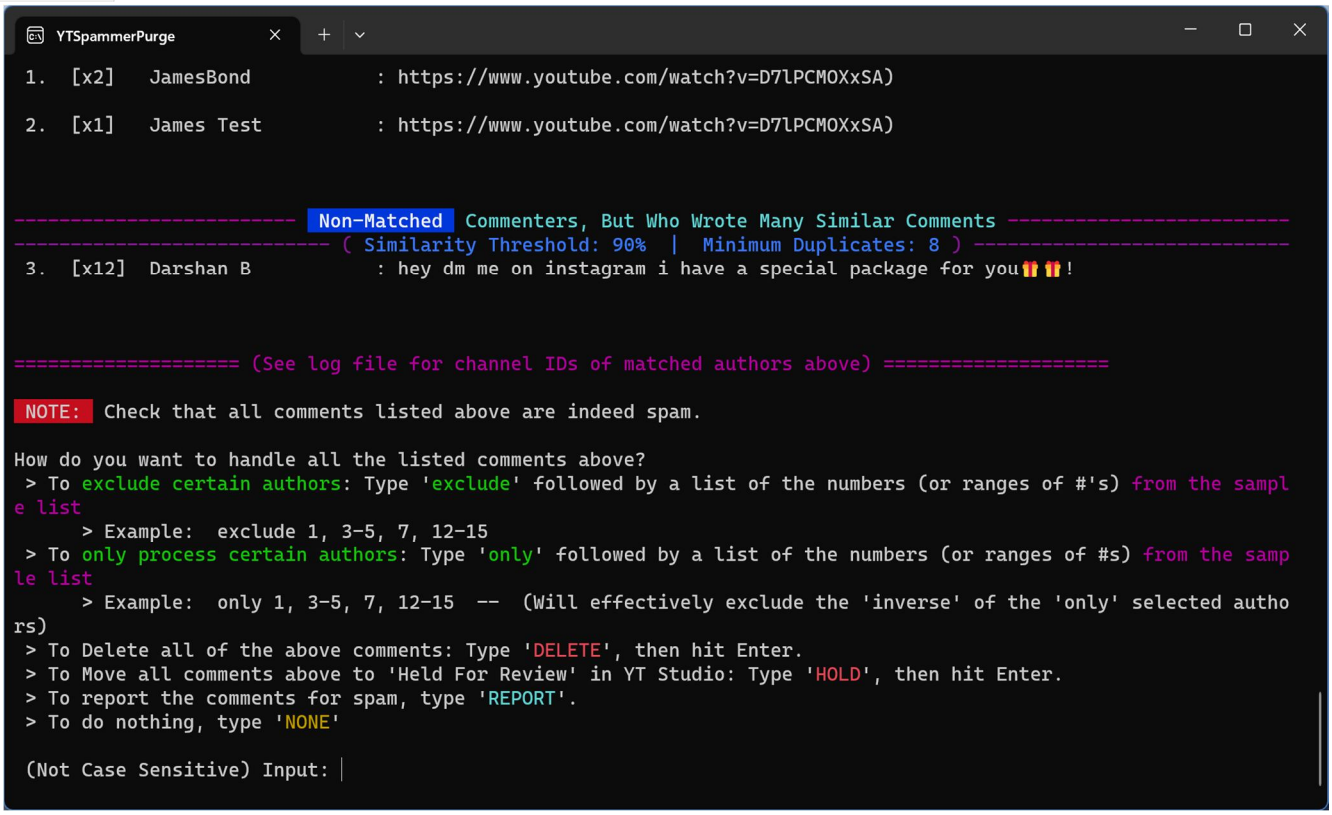


Fig 8. Screen Shot of Detected Spam Comments by our Application

VI. CONCLUSION

In this paper, we have proposed a novel method for detecting and deleting spam comments on YouTube using natural language processing techniques such as Levenshtein distance and the Rapid Fuzz algorithm. We have used YouTube Data API v3 to collect and analyze comments from various popular videos and applied our method to filter out spam comments based on their similarity with previous comments and their content. We have evaluated our method using various metrics such as accuracy, precision, recall, and F1-score and compared it with existing methods such as Support Vector Machine, K-Nearest Neighbour, and Ensemble Machine Learning Model. Our results show that our method outperforms the existing methods in terms of accuracy and F1-score and can effectively detect and delete spam comments on YouTube. We believe that our method can be useful for YouTube users and creators who want to maintain a clean and engaging comment section for their videos. We also suggest some future directions for improving our methods such as incorporating sentiment analysis, topic modelling, and deep learning techniques. Thus, after carrying out the necessary research we try to create a program that will help the YouTube community to identify the spam comments on their channels and hence help them to detect and delete using our tool.

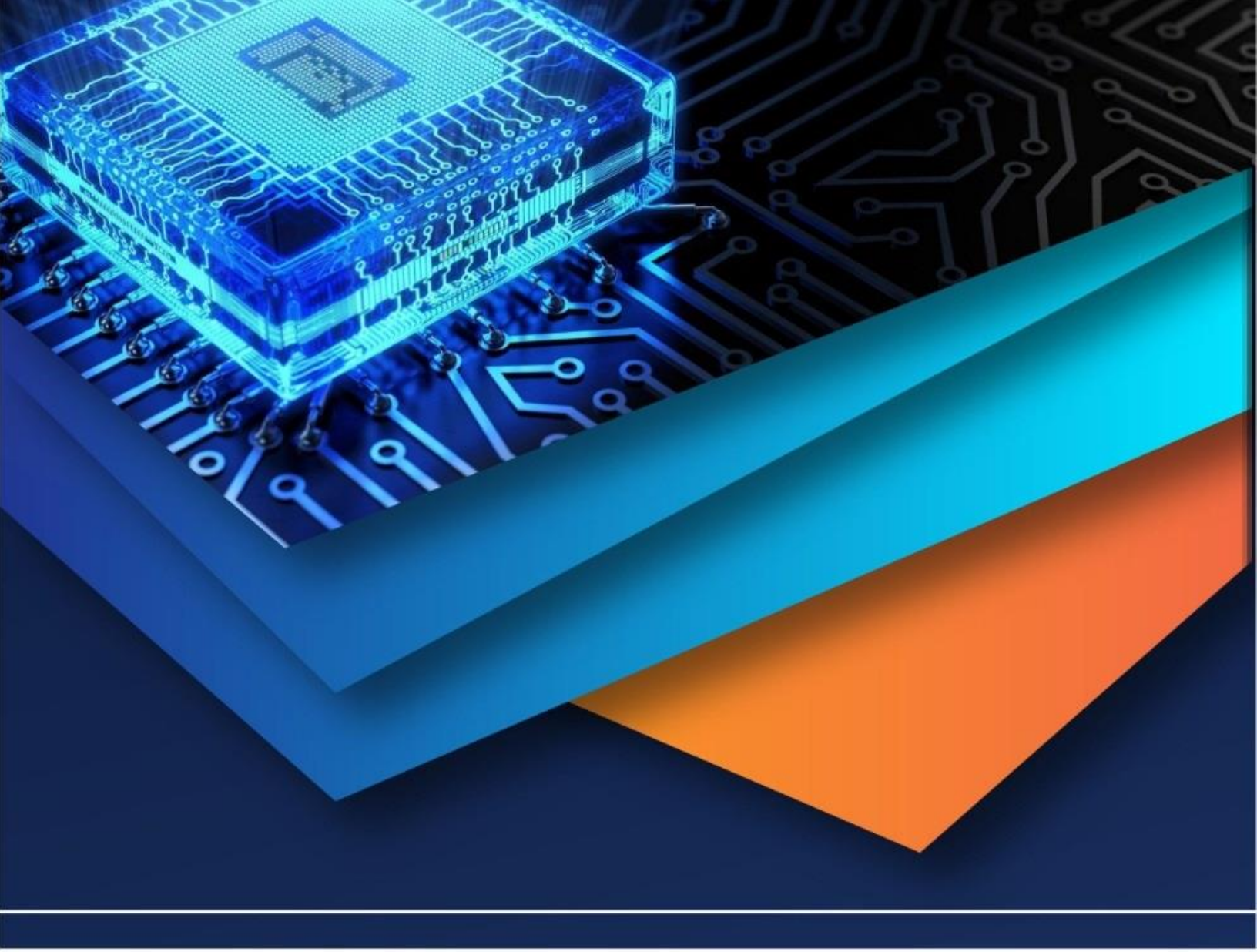
VII. ACKNOWLEDGMENT

We would like to sincerely thank Dhole Patil College of Engineering for their assistance and leadership during the project. Additionally, we would like to thank Prof. Chetana Patil for serving as our project guide and devoting her time and effort to assisting us in finishing this outstanding project. We also like to convey our appreciation to all the writers whose books and academic papers we used as references for this project which made vital contributions crucial to our success.

REFERENCES

[1] S. Aiyar and N. P. Shetty, “YouTube spam comment discovery" Proc. Computer. Sci., vol. 132, pp. 174—182, Jan. 2018, doi:10.1016/j.procs.2018.05.181.
[2] Alberto, J. V. Lochter, and T. A. Almeida, "YouTube Spam: comment spam filtering on Youtube videos" in Proc. IEEE 14th Int. Conf. Mach. Learn.ppl. (ICML4), Dec. 2015, pp. 138-143, doi: 10.1109/1CMLA.2015.37.
[3] R. K. Das, S. S. dash, K. Das "Detection of spam in YouTube comment section using different classification algorithms" in Advanced Computing and Intelligent Engineering, 2020, pp. 201—214, doi:10.1007%2F978- 981-15-1081-6 17.

[4] A. O. Abdullah, M. A. Ali, M. Karabatak, and A. Sengur, "A comparative alysis of common Youtube comment spam filtering methods" in Digit. Forensic Secur. (ISDFS), Mar. 2018, pp. 1—5, doi:10.1109/1SDFS.2018.8355315.
[5] S. Jain and D. M. Patel “Analyzing User Comments of self-learn videos from YouTube platform Using ML Algorithms”.
[6] P. Bansal “Detection of Offensive YouTube Comments, Comparison of Deep Learning Techniques”. Available: https://core.ac.uk/reader/301313034
[7] L. song, R. Y K. Lau, R. C.-W. Kwok, K. Mirkovskl, and w Dou, "Who are the spoilers in social media marketing? Semantics for social spam detection" Electron. Commerce Res., vol. 17, no. 1, pp. 51-81, Mar. 2017, doi: 10.100%10660-016-9244-5.
[8] R Abinaya ,Bertilla Niveda E ,P Naveen “Spam Detection On Social Media Platforms” IEEE 7th International Conference on Smart Structures and Systems ICSSS 2020 doi: 978-1-7281-7223-1/20/ICSSS
[9] Hayoung Oh “Spam Comment detection technique with Cascaded Ensemble ML Model” doi: 10.1109/ACCESS.2021.3121508
[10] A. Kantchelian, J. Ma, L. Huang, S. Afroz “Robust detection of comment spam using entropy rate" in Proc.th ACM Workshop Secur. Artif Intell. (AISec), 2012, pp. 59—70, doi: 10.1145/2381896.2381907.
[11] A. Madden, I. Ruthven "A classification technique of content analyses on Youtube video comments" J. Documentation, vol. 69, no. 5, pp. 693-714, Sep. 2013, doi: 10.1108/JD-06-2012-0078.
[12] N. M. Samsudin, C. F. B. Mohd Foozy, N. Alias, P. Shamala, N. E Othman “YouTube spam detection using Naïve Bayes and logistic regression algorithms" Indonesian J. Electr. Eng. Comput. sci., vol. 14, no. 3, p. 1508, Jun. 2019, doi: 10.11591/ijeecs.v14.i3.pp1508-1517.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)